

## Bioinformatics & Biological Computing Unit



### Human Genome Databases on the Internet (alphabetically ordered)

---

[Click here](#) to view a list of database names only

---

- BCM HGSC, The Baylor College of Medicine Human Genome Sequencing Center.
- BIC, Breast Cancer Information Core.
  - A breast cancer mutation database, for the molecular biologist investigating inherited breast cancer.
- BodyMap, Anatomical Expression Database of Human Genes.
  - BodyMap is a data bank of expression information of human genes, novel or known, in various tissues or cell types. It is created by random sequencing of clones in 3'-directed cDNA libraries. Since these libraries were made so as to faithfully represent the composition of material RNA, redundancy of the same sequence reflects the quantitative aspect of gene expression in various tissues.
- CAD, Chromosome Abnormality Databases.
  - These databases collect records of chromosome abnormalities from over 70% of the laboratories performing routine cytogenetic analysis in the UK. The records contain the karyotype of the sample, the primary reason for referral, the type of tissue sampled and the presence of a cell line/stored tissue.
- CEPH Homepage.
- CEPH genotype database.
  - The Centre d'Etude du Polymorphisme Humain (CEPH) maintains a database of genotypes for all genetic markers that have been tested in the reference families for linkage mapping of the human chromosomes.
  - FTP site (mirror).
- CEPH-Genethon integrated map, A YAC contig map of the human genome.
  - FTP site (mirror).
- CHLC, the Cooperative Human Linkage Center.
  - Genetic maps showing the positions of genetic markers; Search by name for information on markers.
  - FTP site (mirror).
- Chromosome Idiograms, UW Pathology.
  - PostScript version of human chromosome idiograms.
- Chromosome-specific databases.
- CpGIsle, The Human CpG Island database.
  - The CpG island database deals at present with human genes appearing in major releases of the EMBL nucleotide sequence database but it is hoped that in the future it will include islands from other mammalian species. The present release is the third one and is derived from EMBL release 40.
  - FTP site (mirror).
- dbEST, Database of Expressed Sequence Tags (NCBI).
  - The database contains sequence data and other information on "single-pass" cDNA sequences (Expressed Sequence Tags).

- Disease Gens and ESTs. Representation of Positionally-Cloned Genes Mutated in Human Disease in dbEST.
- FTP site (mirror).
- dbSTS, Database of Sequence Tagged Sites (NCBI).
  - The database contains sequence and mapping data on short genomic landmark sequences (Sequence Tagged Sites).
  - FTP site (mirror).
- DHMHD, The Dysmorphic Human and Mouse Homology Database.
  - This application consists of three separate databases of human and mouse malformation syndromes together with a database of mouse/human syntenic regions. The mouse and human malformation databases are linked together through the chromosome synteny database. The database can be used to search for human or mouse malformation syndromes in different ways:-
    - ◻ By specifying specific malformations or clinical features, or chromosome locations.
    - ◻ By Homology.
    - ◻ By asking for human syndromes located at a chromosome region syntenic with a specific mouse chromosome region (and vice versa from human to mouse).
- EGAD, The Expressed Gene Anatomy Database (TIGR).
  - EGAD was constructed by extraction and curation of sequences from GenBank to create a non-redundant set of human transcript (HT) sequences. In some cases, HTs were created by splicing together distinct GenBank accessions for each exon in those transcripts, or by splicing exons from a genomic sequence. Each HT sequence is linked to a human gene (HG). An HG may have more than one HT due to alternative splicing. Alternative splice forms of a gene are linked in HG and HT reports.
- GAC, Genome Annotation Consortium.
  - A Distributed Consortium for High-Throughput Analysis and Annotation of Genomes.
  - The Genome Channel.
- GDB, The Genome Database (Israeli node).
  - GDB (original site at John Hopkins University, Baltimore, Maryland).
- GENATLAS (Jean Frezal).
  - The GENATLAS database compiles the information relevant to the mapping efforts of the Human Genome Project. This information is collected from original articles in the literature or from the proceedings of Human Gene Mapping and Single Chromosome Workshops.
- GenBank, The NIH DNA Sequence Database (NCBI).
  - FTP site.
- GeneCards, a database of human genes (WIS).
  - A database integrating information about the functions of human genes and their products, and of biomedical applications based on this knowledge.
- GeneMap'99, Gene Map of the Human Genome (NCBI and mapping consortium).
- Genes and Diseases (NCBI).
- GENETHON - Human Genome Research Centre.
- GenLink (Washington University).
  - GenLink is a database resource for human genetics that is currently under development. GenLink provides linkage mapping information and software tools that facilitate the integration of physical and genetic linkage data to produce unified maps of the human genome. Researchers interested in identifying genes based on map positions should find the resource helpful.
  - GenLink's WWW interface contains: Genotypes Database, Telomere Database, View Human Meiotic Maps, Data Repository and Additional Resources.
- Genography, Human population genetics database.
- Genomes (NCBI Entrez).

- The Genomes division of Entrez was created to provide a practical approach to the handling of complete genomes (large and small) as well as genetic and physical maps.
- GenotypesDB (Washington University).
  - The GenotypesDB makes all genotypic data used in the construction of linkage maps presented in GenLink easily accessible through the WWW. Users may retrieve genotypes through a query interface where, after selection of map type, chromosome, marker(s), and a family, the user views family genotypes in tabular form with a pedigree drawing indicating relationships. These are the actual data used by linkage programs to build the meiotic linkage maps of the human genome.
- GSDB, The Genome Sequence Data Base.
- HCD, The TIGR Human cDNA Database - is now replaced by TIGR Human Gene Index.
- HOVERGEN, Homologous Vertebrate Genes Database.
  - This database allows one to select sets of homologous genes among vertebrate species, and to visualize multiple alignments and phylogenetic trees. Thus HOVERGEN is particularly useful for comparative sequence analysis, phylogeny and molecular evolution studies. More generally, HOVERGEN gives an overall view of what is known about a peculiar gene family.
- HGMD, Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff.
- HotMolecBase, Molecules for Medicine (WIS).
  - This database contains information about proteins and other biologically active molecules that are critical for our understanding of the molecular processes that underly the development of diseases.
- Human Genome Map Search (Pennsylvania University).
  - The database returns all GenBank entries that have been mapped to the specified positions on human chromosomes. It also finds all the non-human homologs of those entries. This query uses map data from the Genome Data Base.
- Human/Mouse Homology Relationships (NCBI).
  - This is the Seldin/Debry Human/Mouse Homology Map, a table comparing genes in homologous segments of DNA from human and mouse sources, sorted by position in each genome. Most of the loci in this database are genes.
- Human and Mouse Homology, The Jackson Laboratory - Mouse Genome Informatics.
  - Sorted by human chromosome (FTP).
  - Sorted by human symbol (FTP):
- IMGT, The International ImMunoGeneTics Database.
  - an integrated database specialising in Immunoglobulins (Ig), T-cell receptors (TcR) and Major Histocompatibility Complex (MHC) molecules of all vertebrate species. Includes three databases.
- KEGG, Kyoto Encyclopedia of Genes and Genomes. ( Local mirror.)
  - Kyoto Encyclopedia of Genes and Genomes (KEGG) is an attempt to computerize current knowledge of molecular and cellular biology in terms of the information pathways that consist of interacting genes and molecules.
- LDB, The Location Database.
  - LDB is a database for constructing fully integrated genetic and physical maps. The site includes the ldb program which generates an integrated map from partial maps of physical, genetic, regional, mouse homology and cytogenetic data..
  - FTP site (mirror).
- LENS, Linking ESTs and their associated Name Space.
  - The LENS (Linking ESTs and their associated Name Space) database links and resolves the names and identifiers of clones and ESTs generated in the I.M.A.G.E. Consortium/WashU/Merck EST project. The name space includes library and clone IDs and names from I.M.A.G.E. Consortium, EST sequence IDs from Washington University,

- sequence entry accession numbers from dbEST/NCBI, and library and clone IDs from GDB. LENS allows for querying of I.M.A.G.E. Consortium data via all the different IDs. The I.M.A.G.E. Consortium clones and sequences cDNAs for the public domain.
- Marshfield Center for Medical Genetics. Contains Linkage Mapping Information.
  - MIPS - The Human Genome Analysis Project.
    - Computable, indexed sets of entities related to human genes .
    - A functional classification of human proteins based on systematic homology and pattern analysis.
  - Mutation Database Website (R. Cotton).
  - Mutation Databases, Locus-Specific (list generated by HGMD, Cardiff).
  - OMIM, Online Mendelian Inheritance in Man.
    - This database is a catalog of human genes and genetic disorders.
    - FTP site: the complete text.
    - OMIM Gene Map - Chromosome 17.
    - FTP site: OMIM Gene map.
    - FTP site: OMIM Gene map key.
  - PMD, Protein Mutant Database (GenomeNet, Osaka).
    - A literature-based database that contains "natural" and "artificial" proteins mutations. The database is not human-specific.
  - REBASE, the Restriction Enzyme Database.
  - REPBASE, a database of repetitive sequences.
    - FTP site (mirror).
  - RMC, Resource for Molecular Cytogenetics (Lawrence Berkeley National Laboratory/ Univ. of California, San Francisco).
    - The LBL/UCSF Resource for Molecular Cytogenetics is generating physically mapped probes for use in in situ hybridization. This database provides location of probes in FLpter units (fractional location along the chromosome relative to the end of the short-arm telomere) as well as in chromosomal bands. FLpter is defined as the fractional location along the chromosome relative to the end of the short-arm telomere.
    - FTP site.
  - Resources for Human Molecular Cytogenetics (Bari, Italy). This site includes:
    - Partial Chromosome Paints (PCP) derived for somatic cell hybrids, with images. DNA samples from these hybrids are free available to non-profit Institutions
    - FISH data of about 450 YACs (from CEPH), with images.
  - RHdb, The Radiation Hybrid Database (EBI).
    - An archive of raw data with links to other related databases.
    - FTP site (mirror).
  - RLDB, The Reference Library DataBase (ICRF).
    - RLDB (now RZPD) is a database for the storage of data concerning genomic libraries, the clones that make up these libraries, the filters on which the clones are arrayed, and the probes hybridised to these filters.
    - FTP site.
  - Sanger Centre - Human Genome Project.
  - Sanger Centre - Human Radiation Hybrid Mapping.
  - SHGC, Stanford Human Genome Center.
    - The two primary research goals of the Stanford Human Genome Center are the construction of high resolution Radiation Hybrid maps of the human genome and the sequencing of large, contiguous genomic regions.
  - Single Nucleotide Polymorphism, a Database of Single Nucleotide Polymorphisms (NCBI).
  - STS Data Release (Stanford Human Genome Center).
    - The Stanford Human Genome Center is involved in an international effort to build radiation

- hybrid maps of the entire genome. This is a list of Stanford produced STSs which have been uniquely assigned to a chromosome by somatic cell panels (Correll panels 1 and 2). STSs were derived from Genomic DNA and from cDNA.
- Swiss-Prot, Annotated protein sequence database.
    - SWISS-PROT is a curated protein sequence database which strives to provide a high level of annotations (such as the description of the function of a protein, its domains structure, post-translational modifications, variants, etc), a minimal level of redundancy and high level of integration with other databases.
    - FTP site.
  - TelDB, Telomere Database (Washington University).
    - TelDB is a prototype telomere information center accessible through the WWW. TelDB provides new information about telomeric and subtelomeric regions of chromosomes and telomere literature references, including a user interface to the searchable telomere literature database.
    - FTP Site.
  - TIGR Human Gene Index.
    - The TIGR Human Gene Index (HGI) integrates research results from international human gene research projects, including research at TIGR. The goal of this project is to ultimately represent a non-redundant view of all human genes and data on their expression patterns, cellular roles, functions, and evolutionary relationships. The database will also include links to genomic sequences, mapping data, 3D structures, and literature references.
  - GenLoc (UDB), the Unified Database for Human Genome Mapping (Weizmann Institute).
    - UDB presents a Megabase-scale integrated map for each human chromosome, containing genes, Unigene EST-clusters and PCR-amplifiable markers.
  - UniGene, Unique human gene sequence collection (NCBI).
    - This database groups together sets of GenBank sequences that represent the transcription products of distinct genes. These sequences are being used for transcript mapping in collaboration with several genome mapping centers. Some of the clusters have already been localized to chromosomes, but more detailed mapping map information is not available at this time.
    - FTP site (mirror).
  - Utah: Linkage Maps of the Human Genome.
  - Virgil, a library of links between GDB genes and GenBank human sequences (Frederic Achard, Infobiogen).
  - wEST MAP, The Wellcome Trust Centre for Human Genetics Transcript Maps.
  - Washington University-Merck EST Project.
  - WICGR, Whitehead Institute/MIT Center for Genome Research - An STS-Based Map of the Human Genome.
    - This site contains YAC-contig maps anchored to genetic, radiation hybrid and integrated (RH + genetic) maps; List of all STSs assigned to chromosome; and radiation Hybrid Maps.
    - FTP site .



---

For questions or suggestions please contact: [bioinfo@weizmann.ac.il](mailto:bioinfo@weizmann.ac.il)